Syntactic Analysis of Myanmar Language

Win Win Thant, Tin Myat Htwe, Ni Lar Thein
University of Computer Studies Yangon, Myanmar
winwinthant@gmail.com,nilarthein@gmail.com

Abstract

The task of identifying syntactic components is one of the fundamental tasks in natural language processing. The syntactic structure of every language is organized in terms of subject, object, and other grammatical functions, most of which are familiar from traditional grammatical work. We propose a syntactic analyzer for Myanmar language. Our syntactic analysis has two steps: function tagging and grammatical relation. We identify the function tag sets for Myanmar language and function tagging is a preprocessing step to show grammatical relations of Myanmar sentence. In the task of function tagging, which tags the function of Myanmar sentences with correct segmentation, POS (part-of-speech) tagging and chunking information, we use Naive Bayesian theory to disambiguate the possible function tags of a word. We apply context free grammar (CFG) to find out the grammatical relations. Experiments show that our analysis achieves a good result with simple sentences and complex sentences.

1. Introduction

Syntactic parsing, or syntactic analysis, is an important processing step to many language processing applications such as Anaphora Resolution, Machine Translation, and Question Answering. It is the process of analyzing an input sequence in order to determine its grammatical structure with respect to a given grammar. Syntactic relations operate at word-level with the assumption that input sentences are pre-segmented, POS tagged and chunked.

Syntax is the formal relationships between words of a sentence. It deals with word order, and how the words depend on other words in a sentence.

The natural language processing community is in the strong position of having many available approaches to solving some of its most fundamental problems [5]. Syntactic tags are useful for any application trying to follow the thread of the text –they fine the 'who does what' of each clause, which can be useful to gain information about the situation or to learn more about the behavior of words in the sentence [3].

Myanmar is SOV language. It is also a variable word order language. The free word order feature of Myanmar makes parsing a challenging task. Syntactic analysis is a part of the Myanmar to English machine translation project. If high quality translation is to be achieved, language understanding is a necessity. One problem in Myanmar language processing is the lack of grammatical regularity in the language. This leads to very complex Myanmar grammar in order to obtain satisfactory results, which in term increases the complexity in the parsing process, it is desired that simple grammar is to be used. However, this will cause ambiguities in the parse result.

In our approach, we take the chunk level phrase with the combination of POS tag and its category which is the output of a fully described morphological analyzer, which is very important for agglutinative languages like Myanmar. A small corpus annotated manually serves as training data because the large scale Myanmar Corpus is unavailable at present. The relations of the POS tags are obtained from the corpus. Since the large-scale annotated corpora, such as Penn Treebank, have been built in English, statistical

knowledge extracted from them has been shown to be more and more crucial for natural language disambiguation [4]. As a distinctive language, Myanmar has many characteristics different from English. The use of statistical information efficiently in Myanmar language is still a virgin land waiting to explore.

The rest of the paper is organized as in the followings. Next, in the Related Work section, we analyze previous efforts related to the tasks of function tagging and syntactic analysis. Section 3 includes issues of Myanmar language. Section 4 describes syntactic structure of Myanmar language. Section 5 explains about Grammar of Myanmar Language. Section 6 includes the proposed function tag sets. Section 7 describes about corpus creation. Function tagging model is presented in section 8. Section 9 describes about grammatical relations of Myanmar Sentence. Section 10 explains about experimental results. Finally the conclusion of the paper is presented.

2. Related Work

Blaheta and Johnson [1] addressed the task of function tags assignment. They used a statistical algorithm based on a set of features grouped in trees, rather than chains. The advantage was that features can better contribute to overall performance for cases when several features are sparse. When such features are conditioned in a chain model the sparseness of a feature can have a dilution effect of an ulterior (conditioned) one.

Mihai Lintean and Vasile Rus[11] described the use of two machine learning techniques, naive Bayes and decision trees, to address the task of assigning function tags to nodes in a syntactic parse tree. They used a set of features inspired from Blaheta and Johnson [1]. The set of classes they used in their model corresponds to the set of functional tags in Penn Treebank. To generate the training data, they have considered only nodes with functional tags, ignoring nodes unlabeled with such tags. They trained the classifiers on sections 1-21 from Wall Street Journal (WSJ) part of Penn Treebank and used section 23 to evaluate the generated classifiers.

Yong-uk Park and Hyuk-chul Kwon [18] tried to disambiguate for syntactic analysis system by many dependency rules and segmentation. Segmentation is made during parsing. If two adjacent morphemes have no syntactic relations, their syntactic analyzer makes new segment between these two morphemes, and find out all possible partial parse trees of that segmentation and combine them into complete parse trees. Also they used adjacent-rule and adverb subcategorization to disambiguate of syntactic analysis. Their syntactic analyzer system used morphemes for the basic unit of parsing. They made all possible partial parse trees on each segmentation process, and tried to combine them into complete parse trees.

Mark-Jan Nederhof and Giorgio Satta[8] considered the problem of parsing non-recursive context-free grammars, i.e., context-free grammars that generate finite languages and presented two tabular algorithms for these grammars. They presented their parsing algorithm, based on the CYK (Cocke–Younger–Kasami) algorithm and Earley's alogrithm. As parsing CFG (context-free grammar), they have taken a small hand-written grammar of about 100 rules. They have ordered the input grammars by size, according to the number of nonterminals (or the number of nodes in the forest, following the terminology by Langkilde (2000)).

In our system, we use Naïve Bayesian model for its simplicity and user-friendliness. We apply context-free grammar for grammatical relations because it is easier to maintain and can add new language features. Besides, it can automatically construct efficient syntactic analyzer.

3. Issues of Myanmar Language

A number of issues are affecting the syntactic analysis of Myanmar language.

- Myanmar is a free word order. The sentence can be constructed by placing emphatic phrase at the start of a sentence.
  For example:
  သူသည် သတင်းစာကို ဖတ်သည်။
  (Subject+Object+Verb)

သတင်းစာကို သူ ဖတ်သည်။
(Object+Subject+Verb)

- The subject or object of the sentence can be skipped, and still be a valid sentence.
  For example: သွားသည်။ (verb)
- The Burmese language makes prominent usage of particles, which are untranslatable words that are suffixed or prefixed to words to indicate level of respect, grammatical tense, or mood.
  For example:
  မောင်မောင် **များ** ပထမရလျင် သူ့မိဘများက အံ့ဩ လိမ့်မည်။
  (If Mg Mg wins the first prize, his parents will surprise.)
- In our language, an adjective can specialize before or after a noun unlike other languages.
  For example:
  သူမ သည် **လှပသော** မိန်းကလေးတစ်ယောက် ဖြစ်သည်။ (သို့မဟုတ်)
  သူမ သည် မိန်း**မလှ**လေး တစ်ယောက်ဖြစ်သည်။
  She is a beautiful girl.
- The subject /object can be another sentence, which does not contain subject or object.
  For example: ကလေးများ သစ်ပင် အောက်တွင် ကစားနေသည်ကို ကျွန်တော် မြင်သည်။
  (I see the children who are playing under the tree.)
- The postposition of subject or object can be hidden.
  For example:
  သူ**သည်** သတင်းစာ**ကို** ဖတ်သည်။
  (သို့မဟုတ်) သူ သတင်းစာ ဖတ်သည်။
  (He reads newspaper.)
- The postposition of time or place can be omitted.
  For example:
  သူမကျောင်း **သို့** သွားသည်။
  (သို့မဟုတ်) သူမကျောင်းသွားသည်။
  (She goes to the school.)
- The verb can be hidden in a Myanmar sentence.
  For example:
  သူ့ကမောင်လှပါ။ (He is Mg Hla.)

This will cause a lot of problem during syntactic analysis, and a lot of possible relations will be resulted, if only simple grammar is used.

4. Syntactic Structure of Myanmar Language

It is known that many postpositions can be used in a Myanmar sentence. If the words can be misplaced in a sentence, the sentence can be abnormal. There are two kinds of sentence as a sentence construction. They are simple sentence (SS) and complex sentence (CS). In simple sentence, other phrases such as object, time, and place can be added between subject and verb. There are two kinds of clause in a complex sentence called independent clause(IC) and dependent clause (DC).There must be at least one independent clause in a sentence. But there can be more than one dependent clause in it. IC contains sentence final particle (sfp) at the end of a sentence [15][17].

SS=IC+sfp
CS=DC...+IC+sfp
IC may be noun phrase or verb or combination of both.
IC=N... (အိမ်ကလူတွေ)
IC=V   (သွား)
IC=N...+V (ဘုရားမှာပန်းနဲ့ဆီမီးလှူ)
DC is the same as IC but it must contain a clause marker (cm) in the end.
DC=N...+cm (ကျောင်းကဆရာ+ပဲ)
DC=V+cm   (ရောက်+ရင်)
DC=N...+V+cm (စိတ်ထား+ဖြ၍+မှ)

5. Grammar of Myanmar Language

Grammar studies the rules behind languages. The aspect of grammar that does not concern meaning directly is called *syntax*. Myanmar (syntax: SOV), because of its use of postposition (wi.Bat), would probably be defined as a "postpositional language", whereas English (syntax: SVO) because of its use of preposition would probably be defined as a "prepositional language".

There are really only two parts of speech in Burmese, the noun and the verb, instead of the usually accepted eight parts (Pe Maung Tin 1956:195). Most Myanmar linguists [6][12][15][17] accepted there are eight parts of speech in Burmese. Burmese nouns and verbs need the help of suffixes or particles to show grammatical relation.

For example:

ကျောင်းသူများ**သာ** ဂုဏ်ထူးရသည်။

သူတို့သည် အတန်းထဲမှာ ရှိ**ကြ**၏။

Burmese is a highly verb-prominent language and that suppression of the subject and omission of personal pronouns in connected text result in a reduced role of nominals. This observation misses the critical role of postposition particles marking sentential arguments and also of the verb itself being so marked. The key to the view of Burmese being structures by nominals is found in the role of the particles. Some particles modify the word's part of speech. Among the most prominent of these is the particle အ, which is prefixed to verbs and adjectives to form nouns or adverbs.There is a wide variety of particles in Burmese.

For example:

သူတို့သည် မန္တ လေးတွင် ၈ ရက် **တိတိ** လည်ခဲ့သည်။

Stewart remarked that "The Grammar of Burmese is almost entirely a matter of the correct use of particles"(Stewart 1956: xi). How one understands the role of the particles is probably a matter of one's purpose.

6. Proposed Function Tag Sets

The label such as subject, object, time, location, etc. is named as function tags. These are conceptually appealing by encoding an event in the format of "who did what to whom, where, when", which provides useful semantic information of the sentences.

In English Penn Treebank, there are 20 function tags. These tags are categorized into four groups such as Grammatical, Form/Function, Topicalisation and Miscellaneous. In Chinese Penn Treebank (CTB), there are 26 function tags. These tags are categorized into five groups. They are Syntactic Label, Semantic Label, Miscellaneous Label, Clause Type and Discrepancy Label.

To the author's knowledge, this is the first attempt to propose the function tags for our Myanmar language. We propose a set of function tags based on the inflecting system and address the question of assigning function tags to Myanmar words. The function tags are mostly identified with word and postpositional marker (PPM) combination.

For example:

မောင်မြ (သည်၊က)တောရွာ (မှာ၊တွင်၊ဝယ်) နေသည်။
(Mg Mya lives at the village.)

သူ (သည်၊က) နံနက် ၈နာရီ (တွင်၊မှာ) ကျောင်း (သို့ ၊ဆီသို့) သွား သည်။
(He goes to school at 8 o'clock in the morning.)

There are 39 function tags.

Function tag for verb phrase
- Active   Active

Function tags for other phrases
- Subject
  Subj   သူ
  PSubj   သူ   SubjP   သည်
- Direct Object
  Obj   ကော်ဖီ
  PObj   ကော်ဖီ   ObjP   ကို
- Indirect Object
  PIobj   မလှ   IobjP   အား
- Place
  Pla   ရန်ကုန်
  PPla   ရန်ကုန်   PlaP   သို့
- Time
  Tim   မနက်
  PTim   မနက်   TimP   တွင်
- Extract
  PExt   ကျောင်းသားများ   ExtP   အနက်
- Simile
  PSim   မင်းသမီး   SimP   ကဲ့သို့
- Compare
  PCom   သူဦးလေး   ComP   နှင့်အတူ
- Own
  POwn   သူ   OwnP   ၏
- Predicative Complement
  Ada   လှသည်

- Subject Complement
  PcomplS သူမသည် **ဆရာမ** ဖြစ်သည်
- Object Complement
  PcomplO ရွှေကို**လက်စွပ်**လုပ်သည်
  PPcomplO တွန့်တွန့် PcomplOP ဟု
- Use
  PUse တုတ် UseP ဖြင့်
- Cause
  PCau မိုး CauP ကြောင့်
- Aim
  PAim အမေ့ AimP အတွက်
- Conjunction
  CCS လျှင်
  CCM ထို့ကြောင့်
  CCC နှင့်
  CCP ကို
  CCA မည့်

7. Corpus Creation

Corpus is a large and structured set of texts. It is used to do statistical analysis, checking occurrences or validating linguistic rules on a specific universe. Besides, it is a fundamental basis of many researches in Natural Language Processing (NLP). Building of the corpus will be helpful for development NLP tools (such as grammar rules, spelling checking, etc). However, there are very few creations and researches of corpora in Myanmar, comparing to other language such as English.

Our corpus is to be built manually. We extend the POS tagged corpus that is proposed in [13]. The chunk and function tags are manually added to the POS tagged corpus. The corpus contains about 3000 sentences with average word length 15. All sentences are collected from Myanmar textbook of middle school and Myanmar grammar books. They are simple sentences and complex sentences. Manually annotated corpora are valuable but scarce resources. The corpus data will be annotated only up to the sentence level in order to be in the same format for all Myanmar languages. The corpus

size is bigger and bigger because the tested sentences are automatically added to the corpus. VC@Active[မိုးရွာ/verb.common]#CC@CCS[လျှင်/cc.sent] #NC@Subj[ကလေး/noun.person,များ/part.number]#NC @PPla[လမ်း/noun.location]#PPC@PlaP [ပေါ်တွင်/ ppm. place]#NC@Obj[ဘောလုံး/noun.objects]#VC@Active [ကန်ကြ/verb.common]# SFC@Null[သည်/sf]။

Figure 1. A sentence in the corpus

8. Function Tagging Model

We model the problem of assigning function tags as a classification problem. Classifiers are programs that assign a class from a predefined set of classes to an instance based on the values of attributes used to describe the instance. We define a set of linguistically motivated features based on which we characterize the instances. We automatically generate instances from our tagged corpus and then use them to derive Naive Bayesian classifier as solutions to the function tags assignment problem [9].

8.1 Naïve Bayesian Classifier

Before one can build naive Bayesian based classifier, one needs to collect training data [11]. The training data is a set of problem instances. Each instance consists of values for each of the defined features of the underlying model and the corresponding class, i.e. function tag in our case. The development of a naive Bayesian classifier involves learning how much each function tag should be trusted for the decisions it makes. Naive Bayesian classifiers are well-matched to the function tagging problem. The Naïve Bayesian classifier is a term in Bayesian statistics dealing with a simple probabilistic classifier based on applying Bayes' theorem with strong (naïve) independence assumptions. It assumes independence among input features. Therefore, given an input vector, its target class can be found by choosing the one with the highest posterior probability.The probability model for a classifier is a conditional model.

$$P(c_k|x_1, x_2, \ldots, x_i) = P(c_k) * P(x_1, x_2, \ldots, x_i|c_k)$$

Let X=x_1, x_2, x_3, … (x_i, i >=1 and X are features)
C=c_1, c_2, c_3, … (c_k, k>=1 and C are classes)
P (c_k|x_1, x_2, … , xi) is referred to as the posterior probability
P (c_k) as the prior probability
P(x_1, x_2,…,x_i|c_k) as the log likelihood

8.2 Function Tagging by using Naïve Bayes Theory

The label such as subject, object, time, etc. is named as function tags. By function, we mean that action or state which a sentence describes. We investigate the application of the Naive Bayes method to Myanmar function tagging.

Each proposed function tag is regarded as a class and the task is to find what class/tag a given word in a sentence belongs to from a set of predefined classes/tags.

A feature is a tag word with category.

- For example:
  သူ့(**pronoun.possessive**)မှာ(**ppm.subj**)အိမ်(**noun.building**)ရှိသည်(**verb.common**)။

Some are same POS tag and same word but different categories.

- For example:
  မောင်လှ(noun.person)နှင့်(**ppm.compare**)
  တုတ်(noun.objects) နှင့်(**ppm.use**)

A class is a one of the proposed function tags.

- For example:
  သူ(**PSubj**)သည်(**SubjP**)ကျောင်း(**PPla**)သို့(**PlaP**)
  သွားသည်(**Active**)။

  သူ(**Subj**)ကျောင်း(**Pla**)သွားသည် (**Active**)။

Same word may have different function tags.

- For example:
  စားပွဲ(PPla) ပေါ်တွင်(PlaP) **စာအုပ်(Subj)** ရှိသည်
  (Active)။

  သူ(PSubj) သည်(SubjP) **စာအုပ်(PObj)** ကို(ObjP)
  ကျောင်း(PPla) တွင်(PlaP) ထားခဲ့သည်(Active)။

There are many chunks in a sentence such as NC (noun chunk), PPC (postposition chunk), AC (adjective chunk), RC (adverb chunk), CC (conjunction chunk), PC (particle chunk) and VC (verb chunk).

We take the tag word with category of each chunk in the sentence. Some chunks contain one or more POS tag with category.

For example:
NC[ခွေး/noun.animals,တစ်/part.number,ကောင်/particle.type]
We select one of the POS tag (noun.animals) with respect to the chunk.

There are many possible function tags (ft_1, ft_2…ft_i) for each POS tag (pt_1, pt_2…pt_j) with category. These possible tags are retrieved from the training corpus by using the following equation that is prior probability.

P (ft_i|pt_j) = C (ft_i,pt_j) / C (pt_j)   (1)

We calculate the probability between next function tags (nt_1, nt_2….nt_k) and previous possible tags by using the following equation that is log likelihood.

P (nt_k|pt_j) = C (nt_k,pt_j) / C (pt_j)   (2)

We multiply the probabilities from two equations and choose the function tag with the largest number as the posterior probability.

Technically, the task of function tags assignment is to generate a sentence that has correct function tags attached to certain words.

9. Grammatical Relations of Myanmar Sentence

The LANGUAGE defined by a CFG (context-free grammar) is the set of strings derivable from the start symbol S (for Sentence). The core of a CFG grammar is a set of production rules that replaces single variables with strings of variables and symbols. The grammar generates all strings that, starting with a special start variable, can be obtained by applying the production rules until no variables remain. A CFG is usually thought in two ways: a device for generating sentences, or a device if assigning a structure to a given sentence. We use CFG for grammatical relations of function tags.

A CFG is a 4-tuple <N,Σ,P,S> consisting of
- A set of non-terminal symbols N
- A set of terminal symbols Σ
- A set of productions P
  - A-> α
  - A is a non-terminal
  - α is a string of symbols from the infinite set of strings (ΣU N)*

- A designated start symbol S

Example grammar
- SENT   -> SUBJ OBJ PLACE
            VERB | SUBJ VERB
- SUBJ   -> PSUBJ SUBJP
- OBJ   -> POBJ OBJP
- PLACE   -> PPLA PLAP
- VERB   -> ACTIVE

Our description of the syntactic analysis process refers to the example in Fig., which illustrates the sentence "I read the book that is given by my father" ("အဖေပေးသောစာအုပ်ကိုကျွန် တော်ဖတ်သည်။"). This sentence is represented as a sequence of word-tags as "N V CC N PPC PRON V" .It is described as a sequence of chunk as "NC VC CC NC PPC NC VC SFC" and the sentence structure (Sentence) contains separate constituents for the object sentence (Obj-sent) and independent sentence (I-sent), which contains other phrases. Note that this parse tree has had some constituents conflated to comply with the constraint that there be only one constituent per word.

(a) NC [အဖေ/noun.person] # VC [ပေး/verb.common] # CC [သော/cc.adj] # NC [စာအုပ်/noun.objects] # PPC [ကို/ppm.obj] # NC [ကျွန်တော်/pronoun.person] # VC [ဖတ်/verb.common] # SFC [သည်/sf]။

(b) Subj[အဖေ]#Active[ပေး]#CCA[သော]#PObj[စaA ုပ်]# ObjP[ကို]#Subj[ကျွန်တော်]#Active[ဖတ်သည်]။
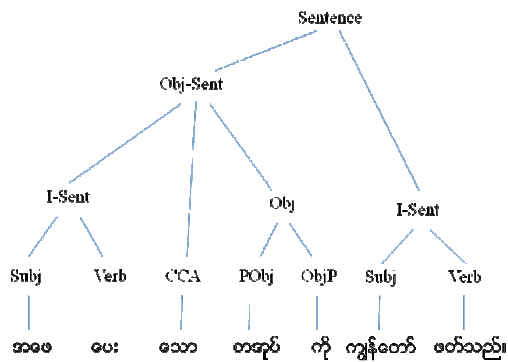
(c)

Figure 2. An overview of the syntactic analysis of the sentence (a) The tagged and chunk sentence (b) The sentence with function tags (c) The parse tree with function tags

10. Error Analysis and Experimental Results

Some errors are observed in error analysis. One typical error is caused by the lack of postpositional markers in a sentence. For example, Pla and Obj are sometimes mistagged when some sentences omit the PPM. The other mistake occurs between Subj and PcomplS. PcomplS are often misidentified as Subj because both tags are especially placed before verb phrase.

Our system is evaluated on different number of sentences collecting from Myanmar textbook of middle school and Myanmar websites. All sentences can be further classified as two sets. One is simple sentence set, in which every sentence has no more than 15 words. The other is complex sentence set, in which every sentence has more than 15 words. In complex sentences, they can be further classified as three groups. They are sentences which are combined by 2 clauses (DC+IC), 3 clauses (DC+DC+IC) and 4 clauses (DC+DC+DC+IC). Therefore, we will obtain complete knowledge about the performance of the syntactic analysis by the comparison of it on these two types of sentences. Table shows the distribution data of simple and complex sentences.

Table 1. Performance of syntactic analysis for simple and complex sentences

Sentence Type	Accuracy
Simple Sentences	94.62%
Complex Sentences with 2 clauses	92.38%
Complex Sentences with 3 clauses	95.51%
Complex Sentences with 4 clauses	96.82%

We found that the accuracy of complex sentences is higher than the simple sentences because clauses in most of the complex sentences form simpler and shorter length than a complete simple sentence.

11. Conclusion

In this paper, we proposed a set of function tags for Myanmar language and investigate the function tag of the word depending on the sentence structure of Myanmar language. We used Naïve Bayesian technique for the task of assigning function tags. Function tags have in the past not been very well studied or exploited. For grammatical relations of the function tags, we use context free grammar. The parse tree can be built by using function tags.

As function tagging is a preprocessing step for grammatical relations, the errors occurred in the task of function tagging affect the relations of the words. The corpus may be balanced because Naïve Bayesian framework probability simply describes uncertainty. The corpus creation is time consuming. The corpus is the resource for the development of Myanmar to English translation system and we expect the corpus to be continually expanded in the future because the tested sentence can be added into the corpus. Because of the lack of prior research on this task, we are unable to compare our results to those of other researchers; but the results do seem promising.

References

[1]Blaheta, D., and Johnson, M. 2000. Assigning function tags to parsed text. In Proceedings of the 1st Annual Meeting of the North American Chapter of the Association for Computational Linguistics, 234–240.

[2]Blaheta, D. 2003. Function tagging. Ph.D. Dissertation, Brown University. Advisor-Eugene Charniak.

[3] Eugene Charniak. 1999. A maximum-entropy inspired parser. Technical Report CS-99-12, Brown University, August.

[4]Eugene Charniak. 1997. Statistical parsing with a context-free grammar and word statistics. In Proceedings of the Fourteenth National Conference on Artificial Intelligence, pages 598-603, Menlo Park.

[5] John C. Henderson and Eric Brill. Exploiting Diversity in Natural Language Processing: Combining Parsers.

[6] Ko Lay, U. 2003 မြန်မာသဒ္ဒါဖွဲ့စည်းပုံ Ph.D. Dissertation, Myanmar Department, University of Educaion.

[7] Leon Versteegen (1999) "The Simple Bayesian Classifier as a Classification Algorithm"

[8] Mark-Jan Nederhof and Giorgio Satta, Parsing Non-Recursive Context-Free Grammars. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL ANNUAL'02), July 7-12, Pages 112-119, Philadelphia, Pennsylvania, USA, 2002.

[9] Michael Collins. 1997. Three generative, lexicalized models for statistical parsing. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, pages 16-23.

[10] Michael Collins. 1996. A New Statistical Parser Based on Bigram Lexical Dependencies. In Proceedings of ACL-96, pp. 184–191.

[11] Mihai Lintean and Vasile Rus. Naive Bayes And Decision Trees For Function Tagging. In Proceedings of the International Conference of the Florida Artificial Intelligence Research Society (FLAIRS) 2007, Key West, FL, May 2007 (in press).

[12]Myanmar Thudda, vol. 1 to 5 in Bur-Myan, Text-book Committee, Basic Edu., Min. of Edu., Myanmar, ca. 1986

[13]P.H.Myint, "Assigning automatically Part-of–Speech tags to build tagged corpus for Myanmar Language", The Fifth Conference on Parallel Soft Computing, Yangon, Myanmar, 2010

[14]Steven Abney. 1991. Parsing by chunks. In Robert Berwick, Steven Abney, and Caroll Tenney, editors, Principle-Based Parsing, pages 257–278. Kluwer Academic Publishers, Dordrecht.

[15]Shwe Pyi Soe, U. 2010. မြန်မာဘာသာစကား Aspects of Myanmar Language

[16]Y. Tsuruoka and K. Tsujii. 2005. Chunk parsing revisited. In Proceedings of the Ninth International Workshop on Parsing Technologies. Vancouver, Canada.

[17]Thaung Lwin, U. 1978 နည်းသစ်မြန်မာသဒ္ဒါ

[18]Yong-uk Park and Hyuk-chul Kwon, Korean Syntactic Analysis using Dependency Rules and Segmentation , Proceedings of the Seventh International Conference on Advanced Language Processing and Web Information Technology(ALPIT2008), Vol.7, pp.59-63, China, July 23-25, 2008